

Fenomenologiczne ujęcie świadomości – eksperymenty myślowe

Tomasz Garbus

2021

1 Wstęp

Thomas Nagel ([5]) pisał, że świadomość jest tym, co czyni problem umysł-ciało trudnym (*intractable*), a nawet beznadziejnym (*hopeless*). Przeciwwstawił on problem umysł-ciało zjawiskom, które nauki szczegółowe dokładnie wyjaśniły dzięki podejściu redukcjonistycznemu. Tak więc pioruny można opisać jako wyładowania elektryczne, wodę jako H_2O , a geny mikrobiolog wytłumaczy w kategoriach struktury kodu DNA.

Oczywiście w wyjaśnianiu umysłu nauka również dokonała niemałych osiągnięć poprzez redukcję faktów wyższego rzędu do faktów na niższym poziomie. Skany fMRI pozwalają naukowcom powiązać konkretne obszary mózgu z odpowiednimi zdolnościami kognitywnymi. Genetyka zachowania stosuje metody statystyczne w badaniu korelacji między uwarunkowaniami genetycznymi a skłonnościami do pewnych zachowań. Przykłady podobnych dokonań są niezliczone, jednak powstały w ten sposób obraz umysłu jest niekompletny. Nawet kiedy wyjaśnimy wszystkie zachowania, zdolności kognitywne, lingwistyczne itd., pozostaje coś jeszcze – subiektywny charakter przeżyć. Słowami Nagela nazwalibyśmy to tym, *jak to jest* być danym umysłem. Chalmers ([2][p. 43]) nazywa to *cechami fenomenalnymi* umysłu (w odróżnieniu od *cech psychologicznych*, które są obiektem badań psychologii oraz kognitywistyki) lub *qualiami*. Bardziej potocznie możemy mówić o świadomości jako o *wewnętrznym filmie* lub *pierwszoosobowym wglądzie*.

Dokonania nauk szczegółowych w wyjaśnianiu konkretnych zachowań czy zdolności poznawczych pomijają ten subiektywny aspekt umysłu. Te same obszary mojego mózgu odpowiadałyby za zdolności językowe, te same geny odpowiadałyby za moją skłonność do wygłaszania pewnych sądów, gdybym w ogóle nie miał pierwszoosobowego wglądu w swoje myśli. Nie twierdzę tu, że świadomość nie ma żadnego znaczenia w wyjaśnianiu zachowania (byłoby to stanowisko epifenomenalisty). Chodzi mi raczej o to, że jej subiektywny charakter uniemożliwia wgląd osobom trzecim, a więc uwzględnianie faktów fenomenologicznych w eksperymentach naukowych. Celem tej pracy nie są jednak rozważania nad rolą faktów fenomenologicznych w tłumaczeniu zachowania. Skupię się na najbardziej fundamentalnym problemie, a mianowicie: jaka klasa systemów ma przeżycia świadome?

To pytanie wydaje się równie beznadziejne, jeśli chcemy posłużyć się empirią w poszukiwaniu odpowiedzi. Nie mamy wglądu w umysły innych osób, możemy co najwyżej pytać o ich doznania. Taka metoda wiedzie jednak na manowce: w końcu pies nie odpowie nam, czy ma bogate życie wewnętrzne, za to dobry maszynowy model językowy skwapliwie przytaknie i jeszcze opowie mi, jak bardzo ceni zapach świeżo skoszonej trawy!

Bardziej obiecującą formą budowania bądź to teorii świadomości, bądź prywatnych intuicji, są eksperymenty myślowe. Nie mają one wprawdzie mocy weryfikacji ani refutacji konkretnych teorii, ale pozwalają wydobyć wewnętrzne sprzeczności w założeniach lub ujawnić niewygodne wnioski, do których założenia te mogą prowadzić.

Przedstawię w poniższej pracy przykłady dwóch eksperymentów-argumentów: argument z blaknących qualiów Davida Chalmersa ([2]) oraz Neda Blocka argument przeciwko funkcjonalizmowi ([1]). Po tych dwóch scenariuszach rodem z science-fiction opowiem o udokumentowanym przypadku prawdziwego pacjenta, którego półkule mózgowe zaczęły "żyć własnym życiem". Następnie zmodyfikuję go nieco na potrzeby własnego eksperymentu myślowego. Jego celem będzie przedstawienie klasy systemów, które teoria pretendująca do miana kompletnej teorii świadomości musi wyjaśniać.

2 Głowa pełna homunkulusów

Ned Block ([1]) konstruuje eksperyment myślowy z głową homunkulusów, aby oskarżyć funkcjonalizm o liberalizm w przypisywaniu przeżyć świadomych systemom, które ich nie mają.

Funkcjonalizm to doktryna ontologiczna głosząca, że stany mentalne¹ są determinowane przez relacje kauzalne z bodźcami zmysłowymi, innymi stanami mentalnymi, oraz wytwarzanym zachowaniem. Dla ustalenia uwagi skupmy się na funkcjonalizmie maszynowym, w którym do opisu relacji między wejściami, wewnętrznymi stanami i wyjściami, korzysta się z formalizmu maszyn Turinga. Rozważmy bardzo prostą maszynę Turinga, odpowiadającą automatowi do słodkich napojów:

wejście / stan	S_1	S_2
moneta pięciocentowa	(S_2, \emptyset)	(S_1, Kola)
moneta dziesięciocentowa	(S_1, Kola)	$(S_1, \text{Kola i pięć groszy})$

Tabela 1: Wartości komórek tabeli są krotkami, których pierwszy element określa, do jakiego stanu powinna przejść maszyna, a drugi – co powinna "wypisać" na wyjście.

Podobnie, jak na tym uproszczonym przykładzie, funkcjonalizm maszynowy określa *explicite* wejścia i wyjścia systemu, podczas gdy dopuszcza większą dowolność w realizacji stanów wewnętrznych. Innymi słowy, funkcjonalisty nie interesuje, *czym są* stany S_1 i S_2 , jaka jest ich natura, jak są realizowane; muszą jednak być w odpowiednich relacjach kauzalnych z wejściami, wyjściami, i ze sobą nawzajem.

Wyobraźmy sobie teraz pewnego człowieka, nazwijmy go Fred, który niczym się nie wyróżnia. Przypuśćmy, że Fred jest moim idealnym bliźniakiem, aby mieć punkt odniesienia do prawdziwego człowieka. Możemy z Fredem porozmawiać, uszczypnąć go, upoić go alkoholem czy pograć w piłkę. Ile czasu byśmy z nim nie spędzili, nie zorientujemy się, że coś jest nie tak – przynajmniej do czasu, aż z jakiegoś powodu postanowimy zrobić mu rezonans mózgu. Wtedy dowiemy się, że w jego czaszce, zamiast zwykłego, biologicznego mózgu, jest jama – powiedzmy, plastikowa, w kształcie pokoju bez okien ani drzwi. Jedna ściana pełna jest diod migających w niesamowicie szybkim tempie, druga zabezpieczona jest małutkimi przyciskami. Na trzeciej – wisi tablica z tajemniczym napisem $S_{281292192722232}$ (przynajmniej ten napis udało nam się dostrzec na skanie, jednak on również zmienia się w błyskawicznym tempie). Jakby tego było mało, po pokoju biega całe gono małych ludków, *homunkulusów*, z których każdy wpatruje się w jedną diodę (kilku homunkulusów może obserwować tę samą), a od czasu do czasu pęch wciśnięć guzik na przeciwległej ścianie, zmasać tablicę i napisać nowy tajemniczy ciąg znaków. Widzimy dokąd to zmierza: każdy homunkulus implementuje jeden kwadrat tabeli opisującej pewną maszynę Turinga² (choć tym razem jest ona niewyobrażalnie wielka w porównaniu do macierzy opisującej automat z kół). Diody podłączone są do neuronów zmysłów (*sensory neurons*) Freda, a guzik odpowiada za kontrolę jego motoneuronów (*wyjść*).

Czy Fred ma qualia? Przynajmniej na pierwszy rzut oka jest wątpliwe, by taki system był świadomy.

2.1 Uwagi

Argument z braku qualii opiera się na intuicji, że Fred nie ma równie bogatych przeżyć świadomych jak ja. Innymi słowy – Block chce, abyśmy uznali za bardzo *nieprawdopodobne*, że Fred ma qualia. Obraz jednak męci fakt, że samo istnienie Freda jest bardzo nieprawdopodobne. Weźmy dwa zdania:

- A. *Istnieje mój bliźniak Fred, po którego głowie biegają homunkulusy.*
 B. *Istnieje mój bliźniak Fred, po którego głowie biegają homunkulusy i ma on przeżycia świadome.*

Rzeczywiście, oba wydają się skrajnie niewiarygodne. Gdybym miał ocenić prawdopodobieństwo zdania $p(A)$ (powiedzmy, że ktoś zmusiłby mnie do postawienia wszystkich moich oszczędności u bukmachera na prawdziwość bądź fałszywość zdania A i liczyłbym wartość oczekiwaną zwrotu z zakładu), byłoby ono niezwykle małe (nie umiem podać dokładnej liczby). Zdaniu B przypisałbym jeszcze mniejszą wartość $p(B)$. Jednak, formalizując intuicję, do której odwołuje się Block, to prawdopodobieństwo warunkowe $p(B|A)$ powinno być bardzo małe. A tego wcale nie uważam za oczywiste. Spróbujmy teraz zastąpić homunkulusy czymś mniej groteskowym:

¹Pojęcie stanów mentalnych jest trochę nieostre. Przypuśćmy, że tydzień temu jadłem na obiad racuchy, a dziś rano ktoś stworzył idealną kopię mojego ciała – identyczną co do komórki. Czy mój idealny klon pamięta jedzenie racuchów – czy jest w *stanie mentalnym* pamiętania o jedzeniu racuchów? Na potrzeby tego eksperymentu myślowego utożsamiam stany mentalne z subiektywnymi aspektami przeżyć. Sądzę, że wspomnienie jedzenia racuchów tydzień temu będzie identyczne dla mnie i mojego klona, nawet jeśli on ich nigdy nie jadł.

²Uważny Czytelnik zapyta: jeśli to maszyna Turinga o pełnej mocy, to gdzie jest jej taśma? Nie widzę powodu, aby nieskończona taśma była potrzebna do symulacji myślenia organizmu, który żyje skończony okres czasu. A jeśli zadowala nas taśma skończona, to równie dobrze może jej nie być, bo efektywnie MT ma wtedy taką samą moc, jak automat o skończonym zbiorze stanów (FSA).

- A'. Istnieje mój bliźniak Fred, w którego głowie znajduje się superkomputer.*
B'. Istnieje mój bliźniak Fred, po którego głowie znajduje się superkomputer i ma on przeżycia świadome.

Prawdopodobieństwa obu zdań niechybnie wzrosły, ale czy jest znaczna różnica między $p(B|A)$ a $p(B'|A')$? Moim zdaniem nie. A przecież wielu ludzi jest skłonnych przyznać świadomość (jeszcze nie istniejącej) sztucznej inteligencji – Block sam oskarża niektóre odmiany funkcjonalizmu o szowinizm właśnie dlatego, że negują posiadanie qualiów przez maszyny!

Podsumowując, moim zastrzeżeniem nie jest to, że istnienie Freda jest mało prawdopodobne lub niemożliwe fizycznie. Chodzi mi o to, że Block przyjmuje jako założenie w swoim argumente, że posiadanie przez Freda przeżyć świadomych jest dalece bardziej nieprawdopodobne niż samo jego istnienie. Mój zarzut to jedynie odrzucenie tej przesłanki³.

3 Argument z blaknących qualiów

Steven Pinker ([2]) w swoim krótkim esej *Could a computer ever be conscious?* rysuje niepokojącą wizję: wyobraźmy sobie, że zespół niezwykle precyzyjnych chirurgów wycina z mojego mózgu jeden neuron, a następnie zastępuje go mikroczipem. Czip ten działa identycznie jak usunięty neuron – na te same impulsy elektryczne i chemiczne odpowiada takimi samymi sygnałami, jak oryginalny neuron. Po zabiegu oczywiście nie zauważam żadnej zmiany. Chirurdzy jednak na tym nie poprzestają; podmieniają mi następny neuron na krzemowy czip, potem trzeci, czwarty, aż większość mojego mózgu staje się krzemowa. Jak będę się czuł w miarę postępowania tych zabiegów? Czy będę miał wrażenie, że umieram, czy może w ogóle nie odczuję żadnej różnicy?

O ile w tekście Pinkera pytanie pozostaje otwarte, Tom Cuda ([3]) konstruuje podobny eksperyment myślowy w obronie funkcjonalizmu przed "neuronowym szowinizmem". Ponownie, wyobraźmy sobie moją idealną kopię – Freda, którego będziemy nazywać też $Fred_0$. $Fred_0$, w odróżnieniu od Freda z poprzedniego argumentu, jest identyczny ze mną nie tylko zewnętrznie, ale również wewnątrz czaszki. Niestety nie nacieszysz się on swoją "normalnością" zbyt długo; po pewnym czasie zostaje poddany operacji opisanej przez Pinkera⁴ – w ten sposób $Fred_0$ zmienia się we $Fred_1$. Po paru dniach odpoczynku, $Fred_1$ znów trafia na stół operacyjny; po podmianie kolejnego neuronu na mikroczip, ze stołu wstaje $Fred_2$. Cykl ten powtarza się tak długo, aż powstaje $Fred_n$, posiadający całkowicie krzemowy mózg.

Argument Cudy jest bardziej złożony i porusza również inne własności umysłu oprócz qualiów (stany intencjonalne, sądy), ale w ujęciu fenomenalnym w uproszczeniu mówi on, że (dla danej sekwencji $Fred_0, Fred_1, \dots, Fred_n$) prawdą musi być dokładnie jedno z poniższych zdań:

1. $Fred_n$ ma równie bogate przeżycia świadome, co $Fred_0$.
2. Dla pewnego m t.ż. $0 < m < n$, $Fred_m$ jest całkowicie świadomy, a $Fred_{m+1}$ – zupełnie pozbawiony qualiów.
3. Qualia Freda powoli zanikają.

Przyjrzyjmy się opcji 3. Wiemy, że Fred nie uświadamia sobie (w sensie psychologicznym) zanikających qualiów – przyjęliśmy założenie, że sztuczne neurony działają identycznie jak prawdziwe, a więc wszystkie fakty psychologiczne o $Fredzie_0$ muszą być też prawdziwe o $Fredzie_m$. Musimy więc przyznać, że jeśli świadomość Freda stopniowo zanika, to on tego nie zauważa. W takim wypadku $Fred_m$ myliłby się na temat własnych przeżyć; mówiłby np. "odczuwam silny ból głowy", tak naprawdę go nie czując. Skoro Fred myli się w kwestii swoich qualiów, to dlaczego my nie popełniamy tego samego błędu? Jakim mam podstawy, by uwierzyć dowolnej innej osobie mówiącej "mam silne doznanie czerwieni", a nie wierzyć Fredowi wypowiadającemu to samo zdanie? Oczywiście, wiem z pierwszoosobowego doświadczenia, że biologiczne neurony wywołują qualia. Cuda jednak argumentuje, że wydaje się niewiarygodne, by obecność lub brak qualiów zależały od arbitralnego wyboru

³Block poświęca jeden rozdział pracy uzasadnieniu tej przesłanki (*Is the prima facie doubt merely prima facie?*), jednak większość tego rozdziału stanowią argumenty przeciwko "zdroworozsądkowej" wersji funkcjonalizmu, czyli takiej, gdzie wejścia, wyjścia i stany wewnętrzne systemu opisywane są prostymi pojęciami jak "ból", "ruch ręki", "zapach róży" itd. Jak sam przyznaje, większość tych argumentów nie uderza w psychofunkcjonalizm (który postuluje precyzyjne, naukowe opisy stanów wewnętrznych, wejść i wyjść). A przecież przyjęliśmy już w eksperymencie myślowym psychofunkcjonalistyczne specyfikacje wejść i wyjść (jako pobudzenia neuronów). W publikacji Blocka znajdziemy silne zarzuty również przeciwko psychofunkcjonalizmowi, ale są one niezwiązane z przytoczonym eksperymentem.

⁴Cuda proponuje wymieniać każdy neuron biologiczny na homunkulusa, ale bez straty ogólności możemy zostać przy konwencji mikroczipów.

materiału, z którego stworzony jest dany mózg. Nazywa taką postawę neuronowym szowinizmem i ją odrzuca.

Opcja druga wydaje się jeszcze bardziej nieprawdopodobna – dlaczego świadomość miałaby zaniknąć po podmianie jakiejś arbitralnej liczby m neuronów, ale nie $m-1$? Musimy więc zaakceptować wniosek, że Fred_n ma równie bogate przeżycia, co Fred_0 .

3.1 Uwagi

Argument ten, w bardzo podobnej postaci, omawia David Chalmers ([2]). Zaczyna on podobnie, od przedstawienia procedury stopniowego podmieniania neuronów w jakimś mózgu na krzemowe odpowiedniki; zauważa dodatkowo, że argument ten można uogólnić na inne odpowiedniki funkcjonalnie izomorficzne (na odpowiednio niskim poziomie – np. neuronalnym) z systemem wyjściowym. Chalmers jako przykłady takich odpowiedników przytacza różne wariacje na temat ludności Chin kontrolującej poszczególne neurony poprzez łącza radiowe⁵.

Chciałbym jednak pociągnąć to uogólnienie jeszcze dalej, aby pokazać, jak możemy użyć argumentu z blaknących qualiów jako zarzutu przeciwko argumentowi Blocka. Przyjmijmy dla uproszczenia, że w sekwencji $\text{Fred}_1, \text{Fred}_2, \dots, \text{Fred}_n$, kolejne neurony do wycięcia wybierane są tak, aby tworzyły spójny obszar mózgu (tzn. sąsiadowały ze sobą). Tym razem zamiast podmieniać je na mikroczipy, niech Fred_k w miejsce wyciętych k neuronów dostanie miniaturowy komputer wykonujący program – fizyczną symulację usuniętego podobszaru mózgu. W szczególności Fred_n zamiast mózgu ma komputer, z odpowiednio podłączonymi wejściami i wyjściami, wykonujący precyzyjną fizyczną symulację ludzkiego mózgu. Program tej symulacji jest funkcjonalnie izomorficzny z oryginalnym mózgiem – w końcu każdy neuron, każde połączenie między neuronami, ma swoją reprezentację w tym programie i jest w takich samych relacjach z reprezentacjami innych neuronów.

Zauważmy, że tym komputerem może być opisany przez Blocka system homunkulusów zamkniętych w małym pokoju (w końcu są one zdolne do implementowania dowolnej maszyny Turinga o skończonej taśmie, a więc symulacja fizyki ludzkiego mózgu nie powinna być dla nich problemem). Możemy więc powtórzyć argument z blaknących qualiów, budując całe spektrum systemów pośrednich między mną a moją kopią o zupełnie innej zawartości czaszki. W tym przypadku jednak na drugim końcu spektrum umieściliśmy system, któremu Block odmawiał posiadania qualiów.

Argumenty Cudy/Chalmersa i Blocka stoją więc w bezpośredniej opozycji i prawdziwe musi być jedno z poniższych zdań:

1. Cuda i Chalmers mylą się w twierdzeniu, że niemożliwe są blaknące qualia (tak, aby istota nie zauważyła, że jej qualia zanikają).
2. Block myli się w twierdzeniu, że osoba z umysłem implementowanym przez homunkulusy nie ma przeżyć świadomych, niezależnie od opisu implementowanej przez homunkulusy maszyny Turinga.

4 Split-brain experiments

Roger Penrose [6] przytacza przypadki pacjentów, u których, w ramach leczenia silnej epilepsji, lekarze dokonali zabiegu przecięcia ciała modelowatego, odpowiedzialnego za przekazywanie informacji między lewą a prawą półkulą mózgu. Na pacjentach przeprowadzone zostały następnie eksperymenty psychologiczne.

W pierwszym eksperymencie, umieszczano pacjentów w takim położeniu, by całkowicie oddzielić pola widzenia obojgu oczu – to znaczy, lewe i prawe oko otrzymywały kompletnie różne bodźce wizualne. Przykładowo, lewemu oku pacjenta pokazywano kubek, a prawemu – ołówek. Pacjent wypowiadał słowa "To jest ołówek", co jest w zgodzie z faktem, że u większości ludzi to właśnie lewa półkula mózgu jest odpowiedzialna za funkcje językowe. Za to lewą ręką – kontrolowaną przez prawą półkulę – pacjent wybierał sosjerkę, a nie kartkę papieru, jako obiekt idący w parze z tym, co widzi.

Jeszcze ciekawszych wyników dostarczają [4], którzy analizowali innego pacjenta – nazwijmy go "P.S.", po takiej samej operacji przecięcia ciała modelowatego. Otóż, w odróżnieniu od wcześniej

⁵Chalmers nawiązuje przy tym częściowo do innego wariantu eksperymentu myślowego Blocka, w którym zamiast homunkulusów, maszynę Turinga implementuje ludność Chin. Jest to jednak pewna modyfikacja, bo Block nie przypisuje pojedynczym homunkulusom ani mieszkańcom Chin ról konkretnych neuronów.

badanych pacjentów, u P.S. prawa półkula wykształciła zdolności językowe⁶. Jedno z przeprowadzonych badań polegało na ocenie słów (poprzez wskazanie liczby palcem) w skali od 1 (*good*) do 7 (*bad*) (wybrane rezultaty przedstawione są w tabeli 2). Okazało się, że prawa półkula konsekwentnie

Słowo	L	R	Różnica (R - L)
<i>Nice</i>	1	6	5
<i>War</i>	3	7	4
<i>Kiss</i>	1	2	1
<i>Liz</i> (Imię dziewczyny badanego chłopca)	1	3	2
<i>Sunday</i>	7	5	-2
<i>Love</i>	1	4	3
<i>Hate</i>	5	7	2
...

Tabela 2: Przykładowe oceny przypisywane różnym słowom przez dwie półkule mózgowe badanego chłopca. Prawa półkula konsekwentnie oceniała słowa bardziej negatywnie, jedynym wyjątkiem było słowo *Sunday*. Pełna tabelka u [4].

oceniała słowa znacznie gorzej od lewej. Co ciekawe, w innym eksperymencie, w którym półkule odpowiadały na pytania postaci "*Czy lubisz TV/szkotę/dom/kościół/mamę/seks/Liz/piwo/...?*" poprzez udzielanie odpowiedzi na skali 1-5 (1 = *bardzo lubię*, 5 = *bardzo nie lubię*), różnice były znacznie mniejsze (najczęściej równe 0). Autorzy [4] tłumaczą to tezą, że podczas pierwszego eksperymentu prawa półkula była w "gorszym nastroju". Najbardziej zaskakujących odpowiedzi obie półkule udzieliły na pytanie o wymarzoną pracę: prawa odpowiedziała, że chce być kierowcą rajdowym, a lewa – kreślarzem!

Podstawowe pytanie, jakie możemy sobie zadać w kontekście tych eksperymentów, brzmi: czy w głowie P.S. rzeczywiście żyły dwa odrębne umysły? Naturalna wydaje się odpowiedź twierdząca – przywykliśmy do tego, by traktować rozwinięte zdolności językowe jako przekonujące świadectwo na rzecz świadomości danego podmiotu⁷.

Przyjmując założenie, że obu półkulom mózgowym towarzyszyły odrębne przeżycia świadome, chciałbym wskazać na pewną trudność, jaką ono rodzi w myśleniu o świadomości. Mózg z nienaruszonym ciałem modelowatym oraz mózg, w którym jest ono całkowicie przerwane, nie wyczerpują przestrzeni wszystkich możliwości, a jedynie okupują dwa końce spektrum. Pomiędzy istnieją stany, w których ciało modelowate jest przecięte w $n\%$. Co więcej, w odróżnieniu od przedstawionych wcześniej eksperymentów myślowych, takie przypadki wydają się być całkowicie fizycznie możliwe – jeśli współczesna chirurgia pozwala na całkowite przecięcie ciała modelowatego bez uśmiercania pacjenta, zgadywałbym, że rozcięcie go tylko częściowo również jest możliwe. Przypuśćmy dla uproszczenia, że ilość informacji wymienianej między półkulami (w danej jednostce czasu) maleje o $n\%$, gdy ciało modelowate jest w $n\%$ uszkodzone.

Stajemy więc przed podobnym pytaniem, jak w eksperymencie z blaknących qualiów – co się dzieje w przypadkach pośrednich? Widzę dwie możliwości:

1. Świadomość rozdziela się nagle dla pewnego n – to znaczy, mózg z ciałem modelowatym rozciętym w $(n - \epsilon)\%$ ma jedną, spójną świadomość, a w $n\%$, każda półkula ma odrębne przeżycia świadome.
2. Świadomość rozdziela się stopniowo. W przypadkach pośrednich przeżycia świadome są częściowo współdzielone między półkulami, a częściowo są odrębne.

Opcję pierwszą proponuję odrzucić, gdyż niezależnie, jakie n wybierzemy, trudno wyobrazić sobie uzasadnienie, dlaczego właśnie ta wartość jest progiem, w którym świadomość się rozdziela. Sądę zatem, że druga opcja jest prawdziwa. Widzimy jednak, jaką trudność rodzi to dla fenomenologicznej teorii świadomości – w przypadku opisanych systemów pośrednich, nie wystarczy binarna klasyfikacja "dany system ma/nie ma qualiów", lecz potrzeba też wyjaśnienia, w jaki sposób część qualiów może być wspólna, a część odrębna dla dwóch półkul.

⁶Choć nie była w stanie wypowiadać słów, komunikowała się poprzez wskazywanie liczb palcem czy układanie liter z gry Scrabble w słowa.

⁷Wprawdzie osiągnięcia w maszynowym modelowaniu języka (np. [8]) zmuszają nas do kwestionowania intuicji, że język jest wystarczającym korelatem świadomości. Podobnie możemy sobie wyobrazić, że Putnamem [7], osobę w głębokiej hipnozie mówiącą płynnie po japońsku, choć nigdy nie uczyła się tego języka – nie jest ona świadoma treści, które wygłasza. Jednak w przypadku rzeczywistego pacjenta z biologicznym mózgiem, z jakiegoś powodu jestem bardziej skłonny zaakceptować zdolności językowe jako świadectwo na rzecz świadomości, niż u maszyny lub u konfaktycznej osoby wykreowanej na rzecz eksperymentu myślowego Putnama.

5 Wnioski

Przedstawiłem dwa eksperymenty myślowe broniące przeciwnych stanowisk względem funkcjonalizmu, czyli doktryny głoszącej, że organizacja funkcjonalna (na odpowiednio niskim poziomie) jest tym, co determinuje stany mentalne danego systemu. Rozwinałem też eksperyment myślowy z blaknących qualiów autorstwa Toma Cudy i Davida Chalmersa, aby pokazać, jak można użyć go jako kontrargumentu przeciwko argumentowi *Absent Qualia* Neda Blocka. Nie opowiadam się po żadnej stronie sporu o funkcjonalizm – moim celem było sprowadzenie tych dwóch eksperymentów myślowych do wspólnego mianownika, aby pokazać, że co najwyżej jeden z nich może być słuszny.

Następnie pokazałem na przykładzie *split brain experiments*, że do rozważań o świadomości nie potrzeba wymyślać zupełnie niewiarygodnych istot. Z pewnością fizycznie możliwe mózgi z częściowo uszkodzonym ciałem modelowatym są klasą systemów, co do których wyjątkowo trudno wnioskować o obecności i charakterze przeżyć świadomych.

Literatura

- [1] Ned Block. Troubles with functionalism. *Minnesota Studies in the Philosophy of Science*, 9:261–325, 1978.
- [2] David Chalmers. *Świadomy umysł*. PWN, 2017. tłum. Marcin Miłkowski.
- [3] Tom Cuda. Against neural chauvinism. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 48(1):111–127, 1985.
- [4] Joseph E. Ledoux, Donald H. Wilson, and Michael S. Gazzaniga. A divided mind: Observations on the conscious properties of the separated hemispheres. *Annals of Neurology*, 2(5):417–421, November 1977. Copyright: Copyright 2016 Elsevier B.V., All rights reserved.
- [5] Thomas Nagel. What is it like to be a bat? *Philosophical Review*, 83(October):435–50, 1974.
- [6] Roger Penrose. *The Emperor’s New Mind: Concerning Computers, Minds, and the Laws of Physics*. Viking Penguin, 1990.
- [7] Hilary Putnam. *Reason, Truth and History*. Cambridge University Press, 1981.
- [8] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.